DMDC

Information and Technology for Better Decision Making

2007 Workplace and Equal Opportunity Survey of Reserve Component Members

Statistical Methodology Report

2007 WORKPLACE AND EQUAL OPPORTUNITY SURVEY OF RESERVE COMPONENT MEMBERS: STATISTICAL METHODOLOGY REPORT

Author, Defense Manpower Data Center
James Green, Westat
Michael Wilson, Westat

Acknowledgments

2007 WORKPLACE AND EQUAL OPPORTUNITY SURVEY OF RESERVE COMPONENT MEMBERS

Executive Summary

This report describes sample design, sample selection, weighting, and variance estimation procedures for the *2007 Workplace and Equal Opportunity Survey of Reserve Component Members* (*2007 WEOR*). The first section of this report presents the sample design and sample selection procedures. The second section describes the assignment of disposition codes in preparation for survey weighting. The third section of this report provides an overview of the weighting approach adopted for the *2007 WEOR* then details the specific weighting adjustments calculated for the 2007 WEOR. This section concludes with a comment regarding use of the Taylor series approximation for variance estimation and the variables created during the weighting of the *2007 WEOR* to facilitate use of Taylor series software. Location, completion, and response rates for the full sample and for subgroups are presented in the final section of this report.

The *2007 WEOR* sampling frame consisted of 805,144 records drawn from the March 2007 *Reserve Components Common Personnel Data System* (RCCPDS) *Master File*. Auxiliary information used to develop the frame was obtained from the March 2007 *Unit Identification Code* (UIC) *Address File* and additional administrative files that were compiled prior to the scheduled starting date of the survey field period: the May 2007 and June 2007 *Defense Enrollment Eligibility Reporting System* (DEERS) *Point-in-Time Extracts* (PITE). Individuals were included on the frame based on membership in both the April 2007 update of the RCCPDS file and the May 2007 PITE. Sample members who subsequently became ineligible were identified by comparison to the May 2007 update of the RCCPDS file and the June 2007 PITE. Individuals not identified as ineligible by administrative records (for example, due to illness or incarceration) and those who became ineligible during the period July 1, 2007 through August 23, 2007 were identified by self- or proxy-report.. A stratified, single-stage random sampling design was used and Reserve component members were sampled with equal conditional probabilities and without replacement within each stratum. Stratum level sample sizes were determined by variance constraints imposed on key parameter estimates for specified domains.

2007 WEOR sample weights were created in four steps. In the first, sampled members were classified using initially assigned disposition codes as eligible respondents, eligible nonrespondents, ineligible members, or members with unknown eligibility. The assignment of final disposition codes was a sequential process that drew upon sample selection, data collection, and returned questionnaire information. Final disposition codes resolved eligibility for cases originally with unknown eligibility. In the second step, a base weight, computed as the inverse of probability of selection, was assigned to each sample member. In the third step, base weights were adjusted for nonresponse in two stages. In the first, base weights were adjusted to account for members whose eligibility was not known at the end of data collection. In the second stage, weights were adjusted to account for eligible members who returned incomplete or non-usable questionnaires. In the fourth and last step, the weights were raked to control totals to reduce bias not accounted for in the previous steps.

Response rates are generally used to measure the success and quality of survey administration. Survey location, completion, and response rates are reported in the final section of this report. In reporting these rates, guidelines recommended by the Council of American Survey Research Organizations (CASRO) were followed. The weighted location, completion, and response rates for the *2007 WEOR* were 96%, 34%, and 32%, respectively.

Table of Contents

List of Tables

Table of Contents (Continued)

Page

List of Figures

2007 WORKPLACE AND EQUAL OPPORTUNITY SURVEY OF RESERVE COMPONENT MEMBERS

Introduction

This report describes sampling and weighting methodologies for the *2007 Workplace and Equal Opportunity Survey of Reserve Component Members (2007 WEOR).* The first section describes the design and selection of the sample. The second section details the assignment of disposition codes for the *2007 WEOR.* The third section of this report documents the weighting of *2007 WEOR* survey data and introduces the variance estimation strategy that will be used by DMDC. The final section provides response rates, including location and completion rates for the full sample and for population subgroups. The design for this survey is based on the general SOFS-A design described in Riemer and Kroger (2002). Information about administration of the survey and detailed documentation of the survey datasets will be provided in the *2007 Workplace and Equal Opportunity Survey of Reserve Component Members: Administration, Datasets, and Codebook* (DMDC, 2008).

The *2007 WEOR* was designed to represent Reserve component members, up to and including paygrade O6, who have served at least six months at the scheduled beginning of the survey fielding period. The sampling frame consisted of 805,144 records drawn from the March 2007 *Reserve Components Common Personnel Data System* (RCCPDS) *Master File.* Auxiliary information used to develop the frame was obtained from the March 2007 *Unit Identification Code* (UIC) *Address File* and additional administrative files were compiled prior to the scheduled starting date of the survey field period: the May 2007 and June 2007 *Defense Enrollment Eligibility Reporting System* (DEERS) *Point-in-Time Extracts* (PITE). Individuals were included on the frame based on membership in both the April 2007 update of the RCCPDS file and the May 2007 PITE. Sample members who subsequently became ineligible were identified by comparison to the May 2007 update of the RCCPDS file and the June 2007 PITE. Individuals not identified as ineligible by administrative records (for example, due to illness or incarceration) and those who became ineligible during the period July 1, 2007 through August 23, 2007 were identified by self- or proxy-report.

A stratified, single-stage random sampling design was used and Reserve component members were sampled with equal conditional probabilities and without replacement within each stratum. Stratum level sample sizes were determined by variance constraints imposed on key parameter estimates for specified domains.

The *2007 WEOR* incorporated a methodological component which has implications on the weighting procedure. These implications are discussed in detail below, but, generally speaking data weighting required the following steps. The first step in weighting computes a base weight, the inverse of selection probability for each sampled person. Since the eligibility of some persons is not known due to nonresponse, an adjustment apportions the weights of the members with unknown eligibility among the known eligible and ineligible members in the sample (second step). The third step adjusts the weights of eligible respondents to account for the eligible members who did not respond to the survey. The final step in weighting rakes weights to frame counts from the beginning of the data collection period. This final step

compensates for some changes in the population that occur between the time of sample selection and data collection.

Response rates for the *2007 WEOR* were computed in accordance with the standards defined by the Council of American Survey Research Organizations (CASRO). The presentation of response rates for the full sample and for subgroups is provided in the last section of this report.

Sample Design and Selection

Target Population

The *2007 WEOR* was designed to represent individuals meeting all of the following criteria:

1. Member of the Selected Reserve in Reserve Unit, Active Guard/Reserve (AGR/FTS/AR; Title 10 and Title 32), Individual Mobilization Augmentee (IMA), and Military Technician programs of the Army National Guard (ARNG), US Army Reserve (USAR), US Naval Reserve (USNR), US Marine Corps Reserve (USMCR), Air National Guard (ANG), and US Air Force Reserve (USAFR);

2. At least six months service at the scheduled beginning of the survey fielding period; and

3. Up to and including paygrade O6.

Fielding of the survey began August 24, 2007 and ended on December 5, 2007.

Sampling Frame

The sampling frame consisted of 805,144 records drawn from the March 2007 *Reserve Components Common Personnel Data System* (RCCPDS) *Master File*. Auxiliary information used to develop the frame was obtained from the March 2007 *Unit Identification Code* (UIC) *Address File.* and additional administrative files that were compiled prior to the scheduled starting date of the survey field period: the May 2007 and June 2007 *Defense Enrollment Eligibility Reporting System* (DEERS) *Point-in-Time Extracts* (PITE). Individuals were included on the frame based on membership in both the April 2007 update of the RCCPDS file and the May 2007 PITE. Sample members who subsequently became ineligible were identified by comparison to the May 2007 update of the RCCPDS file and the June 2007 PITE. Individuals not identified as ineligible by administrative records (for example, due to illness or incarceration) and those who became ineligible during the period July 1, 2007 through August 23, 2007 were identified by self- or proxy-report.

Sample Design

The 2007 WEOR used a single-stage stratified design. Five population characteristics defined the stratification dimensions: Race/Ethnicity, Reserve component, paygrade, gender, and Reserve Program. These are the first five variables shown in Table 1. The frame was partitioned into 147 strata, produced by cross-classification of the stratification variables. Levels were collapsed within dimensions; occasionally, dimensions were collapsed. For example, separate strata were defined for American Indians/Alaskan Natives (AIAN), Hawaiian/Pacific Islanders (NHPI), and individuals with Two or More Races, collapsing all other dimensions. Race/Ethnicity and Reserve component were partially collapsed to define three strata for Warrant Officers and seven for Individual Mobilization Augmentees (IMA), both small population subgroups. Similarly, race/ethnicity and paygrade were partially collapsed to define five U.S. Coast Guard Reserve strata. The remainder of the population was stratified by race/ethnicity

(Hispanic, Black, and White), Reserve component, paygrade, and gender, always preserving race/ethnicity and component boundaries.

Within each stratum, individuals were selected with equal probability and without replacement. However, because allocation of the sample was not proportional to the size of the strata, selection probabilities varied among strata, and individuals were not selected with equal probability overall. Nonproportional allocation was used to achieve adequate sample sizes for small subpopulations of analytic interest, the survey reporting domains. These domains included subpopulations defined by the stratification characteristics, as well as others: activation during the past 12 months, and region of residence. The reporting domain variables are shown in Table 1.

Table 1.
Stratification and Key Reporting Domain Variables

Variable	Categories
Race/Ethnic category[*]	White/Unknown Black Hispanic AIAN Asian NHPI Two or More Races
Reserve Component[*]	U.S. Army National Guard (ARNG) U.S. Army Reserve (USAR) U.S. Naval Reserve (USNR) U.S. Marine Corps Reserve (USMCR) Air National Guard (ANG) U.S. Air Force Reserve (USAFR) U.S. Coast Guard Reserve (USCGR)
Paygrade Group 5[*]	E1-E4 E5-E9 W1-W5 O1-O3 O4-O6
Sex[*]	Male/Unknown Female
Reserve Program[*]	IMA All Others
Region of Residence	Unknown West South North
Activation	Activated during the past 12 months Not activated during the past 12 months
Paygrade Group 6	Enlisted Officer Unknown

* Stratification variables

Sample Allocation

The total sample size was based on precision requirements for key reporting domains. Given estimated variable survey costs and anticipated eligibility and response rates, an optimization algorithm determined the minimum-cost allocation that simultaneously satisfied the domain precision requirements. Anticipated eligibility and response rates were based on the *June 2006 Status of Forces Survey of Reserve Component Members.*

The allocation was accomplished using the DMDC Sample Planning Tool, Version 2.1 (Deever and Mason, 2003). This application is based on the method originally developed by J. R. Chromy (1987), and is described in Mason, Kavee, Wheeless, George, Deever, Riemer, and Elig (1996). The Tool defines domain variance equations in terms of unknown stratum sample sizes and user-specified precision constraints. A cost function is defined in terms of the unknown stratum sample sizes and per-unit costs of data collection, editing, and processing. The variance equations are solved simultaneously, subject to the constraints imposed, for the sample sizes that minimize the cost function. Eligibility rates modify the prevalence rates that are components of the variance equations, thus affecting the allocation; response rates inflate the allocation, thus affecting the final sample size.

Although 95 domains had been defined for the *2007 WEOR* allocation, precision constraints were imposed only on those of primary interest. Generally, the precision requirement was that an estimated prevalence rate of 0.5 have a 95 percent confidence interval half-width no greater than 0.05. Constraints were manipulated to produce an allocation that achieved satisfactory precision for the domains of interest at the target sample size of approximately 83,000 (see Table A-1 in A).

The total *2007 WEOR* sample size was 83,097. Sample sizes by Reserve component are shown in Table 2 for the levels of the stratification dimensions.

Table 2.
Sample Size by Stratification Level, Member Service

	Total	ARNG	USAR	USNR	USMCR	ANG	USAFR	USCGR
Sample	83,097	20,629	15,672	12,345	12,313	9,983	9,967	2,188
Race/Ethnicity								
White/Unknown	19,117	7,128	3,025	1,286	3,509	1,614	1,297	1,258
Black	26,534	5,027	4,812	5,224	2,804	3,851	4,715	101
Hispanic	13,701	2,596	2,334	1,749	4,064	1,312	1,449	196
AIAN	6,257	2,814	1,108	1,011	240	703	297	84
Asian	10,781	3,064	2,718	1,437	1,256	1,161	1,135	10
NHPI	2,942	0	1,674	237	184	411	419	17
Two or More Races	3,765	0	0	1,401	256	931	655	522
Gender								
Male/Unknown	65,414	17,016	11,347	9,263	11,572	7,551	6,832	1,833
Female	17,683	3,613	4,325	3,082	741	2,432	3,135	355
Paygrade								
E1-E4/Unknown	39,919	11,958	6,178	5,196	8,988	3,498	3,331	770
E5-E9	28,096	4,718	4,463	5,278	2,743	5,116	4,782	996
W1-W5	1,269	703	449	30	53	0	0	34
O1-O3	7,382	2,416	2,500	709	133	658	715	251
O4-O6	6,431	834	2,082	1,132	396	711	1,139	137
Program								
IMA	3,344	NA	774	31	576	NA	1,963	NA
All other	79,753	20,629	14,898	12,314	11,737	9,983	8,004	2,188

Assigning Disposition Codes for the *2007 WEOR*

Final case disposition codes for data weighting are based on information from various sources, including administrative records, field operations and survey returns. Although no single source of information is both compete and correct; inconsistencies among these sources were resolved and a final disposition variable (SAMP_DC) was created for the purposes of data weighting.

Survey Control System Dispositions

Table 3 provides a description of the disposition codes resulting from survey operations. These include record and reported ineligibles, complete and incomplete responses, and refused, blank, postal nondeliverables (PND), and other nonrespondents.

Table 3.
Description of 2007 WEOR Survey Control System Disposition Codes (SAMP_DC)

SAMP_DC	Sample Cases	Percentage	Sum of Base Weights	Percent of Base Weights
1 - Record ineligible	972	1.17	5,999	0.75
2 - Ineligible – Self/Proxy-Report	108	0.13	1,436	0.18
3 - Ineligible – Survey Self-Report	808	0.97	5,298	0.66
4 - Eligible – Complete Response	23,170	27.88	251,765	31.27
5 - Eligible – Incomplete Response	2,091	2.52	19,187	2.38
8 - Refused/Deployed/Other	651	0.78	8,867	1.10
9 - Blank	468	0.56	4,100	0.51
10 - PND	4,455	5.36	31,432	3.90
11 - Nonrespondent	50,374	60.62	477,059	59.25
Total	83,097	100.00	805,144	100.00

Final Weighting Disposition Codes

For the purpose of weighting, the several survey operations disposition codes were collapsed into five disposition codes. These are: 1) eligible respondents (ER), 2) eligible nonrespondents (ENR), 3) ineligible sample members (IN), 4) nonrespondents with unknown eligibility (UNK), and 7) extraneous sample members. Table 4 provides a crosswalk between survey operations dispositions (SAMP_DC) and weighting disposition codes (STATUS). Table 5 presents weighting disposition codes in a more compact form.

Table 4.

Mapping Survey Control System Disposition Codes to Weighting Disposition Codes

SAMP_DC	STATUS	Sampled Cases	Percentage	Sum of Base Weights	Percent of Base Weights
1 - Record ineligible	7 - Extraneous	972	1.17	5,999	0.75
2 - Ineligible–Self/Proxy-Report	3 - Ineligible (IN)	108	0.13	1,436	0.18
3 - Ineligible–Survey Self-Report	3 - Ineligible (IN)	808	0.97	5,298	0.66
4 - Eligible–Complete Response	1 - Eligible Respondent (ER)	23,170	27.88	251,765	31.27
5 - Eligible–Incomplete Response	2 - Eligible Nonrespondent (ENR)	2,091	2.52	19,187	2.38
8 - Refused/Deployed/Other	2 - Eligible Nonrespondent (ENR)	651	0.78	8,867	1.10
9 - Blank	4 - Nonrespondent - Unknown Eligibility (UNK)	468	0.56	4,100	0.51
10 - PND	4 - Nonrespondent - Unknown Eligibility (UNK)	4,455	5.36	31,432	3.90
11 - Nonrespondent	4 - Nonrespondent - Unknown Eligibility (UNK)	50,374	60.62	477,059	59.25

Table 5.

Weighting Eligibility Disposition (STATUS)

STATUS	Sampled Cases	Percentage of Sampled Cases	Sum of Base Weights	Percent of Base Weights
1 - Eligible Respondent (ER)	23,170	27.88	251,765	31.27
2 - Eligible Nonrespondent (ENR)	2,742	3.30	28,054	3.48
3 - Ineligible (IN)	916	1.10	6,734	0.84
4 - Nonrespondent - Unknown Eligibility (UNK)	55,297	66.55	512,591	63.66
7 - Extraneous	972	1.17	5,999	0.75
Total	83,097	100.00	805,144	100.00

2007 WEOR Weighting Procedures

Overall Approach

The analysis of survey data from complex sample designs requires the use of weights to: 1) compensate for variable sample member probabilities of selection; 2) adjust for differential member response rates; and 3) improve the precision of survey-based estimates (Skinner et al., 1989). To develop each of the three sets of weights for the *2007 WEOR* survey, the following steps were taken.

First, base weights equal to the reciprocal of the probability of selection were assigned to each member selected for the sample. These base weights were adjusted for the 10% and 90% sample/mode assignment. Next, the base weights were adjusted for nonresponse using weighting classes defined by relevant variables available on the March 2007 sampling frame file. Finally, the nonresponse-adjusted weights were raked to population counts from the updated August 2007 Reserve Components Common Personnel Data System (RCCPDS) Master File frame. This last adjustment compensates for changes in the eligible population taking place between the time of sample selection and the beginning of data collection. Details of this weighting methodology are described in this section.

The *2007 WEOR* incorporated an experimental design component in which 10% of the sample was offered Web-only access to the survey instrument and 90% of the sample was offered the more usual Web and paper modes. This experimental component presented weighting implications for the *2007 WEOR*. See Figure 1 for an overview of weighting activities. They required the development of three sets of weights. The first set of weights was constructed to support analyses for the 90% of the sample that was offered both the Web and paper modes of survey administration. This set of weights incorporated nonresponse and raking weighting adjustments. The second set of weights was constructed to support analyses for the 10% of the sample that was offered Web-only access to the survey instrument. This set of weights also incorporated nonresponse and raking weighting adjustments. Analysts can use these two sets of weights to detect any differences in survey estimates that may be attributable to the mode(s) offered. Depending on the results of that analysis, a third set of weights may be used which combines the data from both treatment groups. This third set of weights also incorporated nonresponse and raking weighting adjustments, however the nonresponse adjustments were done separately for the different modes, whereas the raking adjustment were done using all cases. The remainder of this section describes the weighting methodology applied regardless of experimental group.

Figure 1.
Flowchart of Weighting Activities
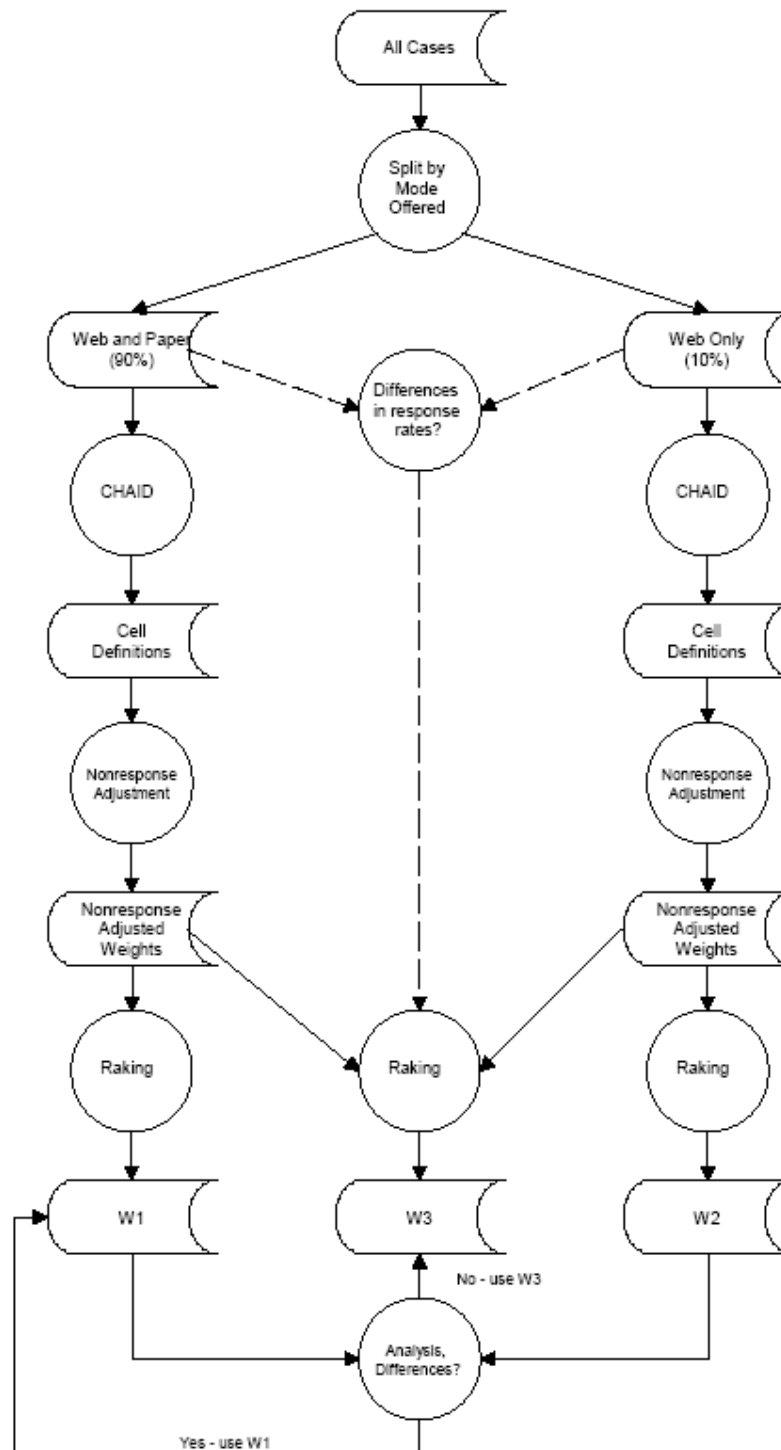
Calculation of Base Weights

The *2007 WEOR* sample was randomly selected without replacement from a stratified frame. As such, the overall probabilities of selection vary by design strata in order to satisfy the precision goals specified by the study. Let U be the frame of the N persons in the population (i.e., Reserve members at the time of sampling). Note that the frame size N includes some persons who were ineligible at the time the survey was conducted because, for example, they had separated from the Reserves. The frame U was partitioned into H non-overlapping strata $U_1,\ldots,U_H$ consisting of N_h persons in each stratum h so that

$$N = \sum_{h=1}^{H} N_h.$$

A simple random sample of size n_h was selected without replacement within each stratum U_h. Given this design, the base weight for the i-th sampled member in stratum h was calculated as:

$$w_{hi} = \frac{N_h}{n_h} \qquad i = 1,\ldots.n_h.$$

For each individual classified in stratum h, the base weight is the ratio of the total number of individuals in the stratum to the stratum-level sample size. The base weight w_{hi} is equal to the reciprocal of the probability of selection and is attached to each sampled person in the data file. Note that n_h is the number of persons initially sampled in stratum h without regard to whether or not the member ultimately participated in the survey.

Nonresponse Weighting Adjustments

Ideally, all the persons in the inference population are eligible to be selected into the sample and all those that are selected participate in the survey. In practice, neither of these conditions occurs. Some of the sampled persons do not respond (unit nonresponse); some sample persons are discovered to be ineligible; and the eligibility status of some persons cannot be determined. If these problems are not addressed, survey estimates will be biased. Nonresponse weight adjustments deal with unknown eligibility and unit nonresponse. Raking is used to account for changes in the distribution of the population between the times of sampling and data collection and increase the precision of survey estimates. The following describes these methodologies in detail.

Unit Nonresponse Adjustments

Unit nonresponse (i.e., complete questionnaire nonresponse) occurs when a sampled member fails to respond for any reason. For example, nonresponse could result from failure to locate the member because of mobility or invalid/incorrect addresses in the frame, or from the unwillingness of some members to participate in the survey. Because the (unweighted) response rate in the survey will be substantially less than 100 percent, adjusting for unit nonresponse is an important step in bias reduction.

To compensate for losses due to nonresponse, we adjusted weights in two stages. The first stage of adjustment accounts for the fact that the eligibility status of some sample persons cannot be determined. The second stage of adjustment compensates for losses due to eligible sample persons who did not complete the questionnaire. At each stage the base weights of usable cases were inflated to account for cases that were unusable. These adjustments were done within classes that cluster persons with similar response rates and other characteristics together. This form of adjustment is referred to as weighting class adjustment; since it adjusts the weighted distribution of respondents across the weighting classes to that of the total sample (Kalton and Kasprzyk, 1989).

One potential drawback to nonresponse adjustment is that it may increase the variability of weights and, thus, increase the sampling variance of some estimates (Kish, 1992). Ideally, the reduction in bias from using a nonresponse adjustment more than compensates for any increase in variance. When the cells contain sufficient cases and the adjustment factors do not become either inordinately large or substantially different from each other, the effect on variance is modest. Very large adjustment factors or factors that are much different from others can occur in cells with high nonresponse rates or small numbers of respondents. To avoid the second situation, cells with few cases were combined to form a new cell with a minimum of 30 cases.

For sample weighting adjustments to be effective in reducing nonresponse biases, it is desirable that the weighting classes be internally homogeneous with respect to response propensity. Equivalently, a criterion for constructing the weighting classes is that the variation in response propensity between the classes be as large as possible without unduly inflating sampling variances. The criteria used to create the classes are described below.

As discussed previously, each sampled member was assigned to an appropriate response-status group (*ER*, *ENR*, *IN*, or *UNK*[1]). At the first stage of weight adjustment, we assumed that the unknowns (Group *UNK*) were distributed among the *ER*, *ENR*, and *IN* categories had it been possible to determine their status. The first-stage nonresponse adjustment factor was calculated within weighting class *c* as:

$$
f_c^{A1} = \begin{cases} \dfrac{\displaystyle\sum_{i \in ER_c} w_i + \sum_{i \in ENR_c} w_i + \sum_{i \in UNK_c} w_i}{\displaystyle\sum_{i \in ER_c} w_i + \sum_{i \in ENR_c} w_i} & \text{If the } i\text{-th sample person classified in weighting class } c \text{ belongs to response group } ER_c, \text{ or } ENR_c. \\[2ex] 1 & \text{If the } i\text{-th sample person in class } c \text{ belongs to eligibility group } IN_c. \\[2ex] & \text{If the } i\text{-th sample person in class } c \text{ is in } UNK_c. \\[2ex] 0 & \end{cases}
$$

The sums in the numerator of f_c^{A1} extend over the following types of persons in class *c*: eligible respondents (*ER*), eligible nonrespondents (*ENR*), and the unknowns (*UNK*). The

---

[1] The response-status groups are defined as ER, eligible respondents, ENR, eligible nonrespondents, IN, ineligible members, and UNK, member with unknown eligibility.

term w_i is the base weight for the i-th sampled person in class c. (As a notational convenience, the subscript h is omitted for the sampling stratum since a class c may extend across strata. However, as described subsequently, the eligibility adjustments and the nonresponse adjustments will be almost always be made using classes that are subdivisions of design strata or the design strata themselves.)

The first nonresponse-adjusted weight w_i^{A1}, for a sample member in class c was computed as:

$$w_i^{A1} = f_c^{A1} w_i$$

Thus, if persons with unknown eligibility accounted for 50 percent of the weight in class c, the weights on the other units are increased by a factor of 2.

The second nonresponse adjustment increases the adjusted weight of eligible respondents to account for eligible nonrespondents. The second-stage nonresponse adjustment factor for class c was computed as:

$$f_c^{A2} = \begin{cases} \dfrac{\displaystyle\sum_{i \in ER_c} w_i^{A1} + \sum_{i \in ENR_c} w_i^{A1}}{\displaystyle\sum_{i \in ER_c} w_i^{A1}} & \text{If the } i\text{-th sample person in weighting class } c \text{ belongs to response group } ER_c. \\[2em] 0 & \text{If the } i\text{-th sample person sampled in weighting class } c \text{ belongs to response group } ENR_c. \\[1em] & \text{If the } i\text{-th sample person in weighting class } c \text{ belongs to response group is in } IN_c. \\[1em] 1 & \end{cases}$$

The first sum in the numerator of f_c^{A2} for eligible respondents extends over the respondents (Group ER) in class c; the second over the eligible nonrespondents (Group ENR) in class c; and w_i^{A1} is the previously adjusted weight of the i-th sample member.

The second nonresponse-adjusted weight w_i^{A2} for the i-th sample member classified in weighting class c was computed as:

$$w_i^{A2} = f_c^{A2} w_i^{A1}.$$

After the two stages of nonresponse adjustment, the weight for a respondent in weighting class c becomes

$$w_i^{A2} = f_c^{A2} f_c^{A1} w_i.$$

13

Note that after the two stages of nonresponse adjustment, the persons with non-zero weight are those in *ER* and *IN*. The members with unknown eligibility (*UNK*) and eligible nonrespondents (*ENR*) have zero weight.

Construction of Weighting Classes

The main objective in constructing weighting classes is to group respondents and nonrespondents with similar characteristics into the same cells. Ideally, the characteristics should be related to both the likelihood of responding to the survey and to values of the data items collected. Each of the characteristics must be available for all initial sample persons in order to create classes. In the *2007 WEOR* sampling strata were used as the starting point for the creation of weighting classes. The sampling strata were created from variables related to survey response propensity and/or important domains in survey topics.

The creation of weighting classes depends in large measure on the number of respondents in the sampling strata. Weighting class will corresponds to sampling stratum when the number of respondents is greater than 30 and smaller than 500. Any stratum with fewer than 30 respondents was combined with another "nearby" stratum to form a new weighting class. When combining strata, the characteristics for Service and race/ethnicity were preserved. These two stratification variables were considered hard boundaries not to be crossed when combining strata. Combining strata defined by main paygrade groups (Enlisted, Warrant Officers and Commissioned Officers) was also avoided whenever possible.

Strata with more than 500 respondents were subdivided into smaller weighting classes. This subdivision into smaller cells was done using a categorical search algorithm called CHAID (Chi-squared Automatic Interaction Detector) (Kass, 1980). CHAID attempts to divide the data set into groups so that response rates between cells are as different as possible. Given a set of categorical predictors of response probabilities, CHAID divides the data set into groups in a stepwise fashion. Through a series of chi-square tests for equality of distributions, CHAID identifies the most important predictor of response rates and splits the data set into categories based upon response rate differences. In subsequent steps, each of those categories is further segmented based on other predictors. Categories of a variable that are not significantly different can be merged together. The merging and splitting continues until no more statistically significant predictors are found or until a user-specified stopping rule is met. The rule imposed for the *2007 WEOR* was that no more than six cells would be formed within large strata and each subdivision would contain at least 30 respondents.

Since the CHAID analysis was carried out separately for each large stratum, we used additional (nonstratification) variables as predictors for nonresponse. Table 6 lists the variables from the administrative record files that were considered for subdividing large strata in addition to the stratification variables.

Table 6.

Member Characteristics Considered for Creation of Nonresponse Weighting Classes Within Strata With 500 or More Respondents (if available)

Description	Values
Race/Ethnicity	White Black Hispanic AIAN Asian NHPI Two or More Races Unknown
Reserve Component	ARNG USAR USNR USMCR ANG USAFR USCGR
Reserve Program	TPU (Reserve Drilling Units) AGR/FTS/AR (Active Guard/Reserve; Title 10, Title 32) Military Technicians IMA (Individual Mobilization Augmentee) Unknown
Paygrade	E1 … E9 W1 … W5 O1 … O6

Table 6. (continued)

Description	Values
Gender	Male Female
Activation Status	Not active in prior 12 months Active in prior 12 months De-activated in prior 11 months
Census Region	U.S. Northeast U.S. South U.S. Midwest U.S. West
Level of Education	Less than High School High School Graduate Some College, but less than a 4-Year Degree 4-Year College Graduate, Graduate School
Age Group	17-18 19-20 21-22 … 63-64 65-66 Unknown
Marital Status	Married Unmarried Unknown

Race/Ethnicity, Reserve component, Reserve Program, paygrade, gender, activation status, Census region, education, age, and marital status, have been identified as predictors of nonresponse in previous DMDC surveys of Reserve component members and so were considered as candidate variables for CHAID processing. The initial assessment of the variables was based on an analysis of response rates for the overall population and by the stratification variable levels of race/ethnicity, Reserve component/Reserve Program and paygrade.

Nonresponse adjustment was done within each weighting class created from the original, combined or split sampling strata. We examined any having unusually large values of the f_c^{A1}, or f_c^{A2} adjustments. When the weighting cells contain sufficient cases and the adjustment factors do not become either inordinately large or substantially different from each other, the effect on the variances is modest. Very large adjustment factors or factors that are much different from others can occur in cells with high response/nonresponse rates or with a small number of respondents. Combining cells with few cases to form new cells with at least 30 respondents was implemented to mitigate large adjustment factors.

Raking Adjustment

The nonresponse-adjusted weights were raked to force sample estimates of numbers of persons to equal known population totals. In the *2007 WEOR* survey, the functions of raking were to:

- Reduce variance;

- Adjust the March 2007 sample to reflect the August 2007 distribution among categories defined by the raking dimensions; and

- Reduce noise in the estimates of mode differences due to the methodological component.

The population or controls were produced using the August 2007 Reserve Components Common Personnel Data System (RCCPDS) *Master File* updated frame. The updated frame reflects any changes in the population between the time of sampling and the beginning of data collection period.

To compute the control totals, we used the variable RELIG defined for all the records on the frame, including both sampled and nonsampled members. The variable RELIG summarizes the eligibility of the member using the March 2007 RCCPDS *Master File* updated frame. The control totals for each raking dimension were computed by counting the eligible members in the matched frames using the member characteristics as of the June frame.

The mechanics of the raking weight adjustment is now summarized. The population is partitioned, based on the first raking dimension, into groups denoted by $U_1, \ldots, U_G$. The groups will be, by definition, mutually exclusive and cover the entire population. Let N_g be the size of U_g, so that $N = \sum_{g=1}^{G} N_g$. The eligible respondents in the sample will also be partitioned into groups $s_1, \ldots, s_G$. The expression for the initial weighting adjustment factor for all the units classified in cell g is

$$\tilde{f}_g^R = \frac{N_g}{\displaystyle\sum_{i \in s_g} w_i^{A2}} \; .$$

The raked weight $\tilde{w}_i^R$ for the *i*-th sample person classified in cell g of the first raking dimension is then computed as:

$$\tilde{w}_i^R = \tilde{f}_g^R w_i^{A2}, i \in s_g$$

A similar adjustment is made after classifying the sample based on the second raking dimension, and so on, for all dimensions included in the raking. Successively adjusting the weights based on all dimensions constitutes the first iteration of the process. The adjustments for

dimensions 2 and greater result in the sum of weights for persons classified by dimension 1 not equaling the control totals for dimension 1. The adjustments for dimensions 1 and beyond are then repeated beginning with the adjusted weights from the first iteration. The iterative process continues until the sum of the weights for each raking dimension is acceptably close to the corresponding control total. Acceptable for *2007 WEOR* is a relative error of less than 1 percent. The final raked weight w_i^R for the *i*-th sample person is then computed as:

$$\tilde{w}_i^R = \tilde{f}_g^R w_i^{A2}, \quad i \in s_g$$

where f_i^R is the product of the iterative adjustments applied to the *i*-th sample person.

For the *2007 WEOR*, we used the raking dimensions presented in Table 7. Tables 8 through 11 present the categories and control totals used for each dimension.

Table 7.
Combinations of Variables Used for Raking Dimensions

Dimension	VARIABLES
DIM1	Reserve Component by Race/Ethnicity
DIM2	Reserve Component by paygrade
DIM3	Reserve Component by gender
DIM4	Reserve Component by Reserve Program

Table 8.
Definition and Control Totals for the First Raking Dimension (DIM1)

DIM1	Description	Control Total
1	ARNG, Hispanic	27,508
2	ARNG, White/Unknown	269,138
3	ARNG, Black	46,572
4	ARNG, AIAN	2,906
5	ARNG, Asian	6,580
6	USAR, Hispanic	22,739
7	USAR, White/Unknown	115,366
8	USAR, Black	42,565
9	USAR, AIAN	1,238
10	USAR, Asian	5,814
11	USAR, NHPI	1,819
12	USNR / USCGR, Hispanic	8,869
13	USNR, White/Unknown	45,096
14	USNR / USCGR, Black	10,779
15	USNR / USMCR / USCGR, AIAN	1,546
16	USNR, Asian	2,670
17	USNR / USMCR, NHPI	478
18	USNR / USMCR, Two or More Races	1,806
19	USMCR, Hispanic	4,735
20	USMCR, White/Unknown	29,304
21	USMCR, Black	2,915
22	USMCR, Asian	1,271
23	ANG, Hispanic	6,507
24	ANG, White/Unknown	85,648
25	ANG, Black	8,863
26	ANG, AIAN	737
27	ANG, Asian	2,377
28	ANG, NHPI	429
29	ANG, Two or More Races	990
30	USAFR, Hispanic	5,356
31	USAFR, White/Unknown	51,994
32	USAFR, Black	11,169
33	USAFR, AIAN	304
34	USAFR, Asian	1,468
35	USAFR, NHPI	432
36	USAFR, Two or More Races	719
37	USCGR, White/Unknown	6,032
38	USCGR, Asian, NHPI, Two or More Races	576
Total		835,315

Table 9.
Definition and Control Totals for the Second Raking Dimension (DIM2)

DIM1	Description	Control Total
1	ARNG, E1-E4/Unknown	176,526
2	ARNG, E5-E9	139,176
3	ARNG, W1-W5	6,683
4	ARNG, O1-O3	18,567
5	ARNG, O4-O6	11,752
6	USAR, E1-E4/Unknown	74,159
7	USAR, E5-E9	79,579
8	USAR, W1-W5	2,711
9	USAR, O1-O3	13,917
10	USAR, O4-O6	19,175
11	USNR, E1-E4/Unknown	18,471
12	USNR, E5-E9	35,226
13	USNR, W1-W5, O1-O3	3,945
14	USNR, O4-O6	11,712
15	USMCR, E1-E4/Unknown	27,042
16	USMCR, E5-E9	8,590
17	USMCR, W1-W5, O1-O6	3,318
18	ANG, E1-E4/Unknown	23,953
19	ANG, E5-E9	67,834
20	ANG, O1-O3	5,088
21	ANG, O4-O6	8,676
22	USAFR, E1-E4/Unknown	14,809
23	USAFR, E5-E9	40,365
24	USAFR, O1-O3	4,516
25	USAFR, O4-O6	11,752
26	USCGR, E1-E4/Unknown	2,898
27	USCGR, E5-E9	3,549
28	USCGR, W1-W5, O1-O6	1,326
Total		835,315

Table 10.
Definition and Control Totals for the Third Raking Dimension (DIM3)

DIM3	Label	Control Total
1	ARNG, Male/Unknown	303,768
2	ARNG, Female	48,936
3	USAR, Male/Unknown	145,240
4	USAR, Female	44,301
5	USNR, Male/Unknown	55,547
6	USNR, Female	13,807
7	USMCR, Male/Unknown, Female	38,950
8	ANG, Male/Unknown	86,484
9	ANG, Female	19,067
10	USAFR, Male/Unknown	53,839
11	USAFR, Female	17,603
12	USCGR, Male/Unknown	6,628
13	USCGR, Female	1,145
Total		835,315

Table 11.
Definition and Control Totals for the Fourth Raking Dimension (DIM4)

DIM4	Label	Control Total
1	ARNG, TPU/Unknown	307,310
2	ARNG, AGR/FTS/AR	25,658
3	ARNG, Military Technicians	19,736
4	USAR, TPU/Unknown	163,414
5	USAR, AGR/FTS/AR	15,356
6	USAR, Military Technicians	6,499
7	USAR, IMA	4,272
8	USNR, TPU/Unknown	56,878
9	USNR, AGR/FTS/AR / IMA	12,476
10	USMCR, TPU/Unknown	34,179
11	USMCR, AGR/FTS/AR	2,205
12	USMCR, IMA	2,566
13	ANG, TPU/Unknown	69,598
14	ANG, AGR/FTS/AR	13,147
15	ANG, Military Technicians	22,806
16	USAFR, TPU/Unknown	48,022
17	USAFR, AGR/FTS/AR	2,401
18	USAFR, Military Technicians	9,073
19	USAFR, IMA	11,946
20	USCGR, TPU/Unknown	7,773
Total		835,315

Variance Estimation

A widely used method of estimating variance for complex surveys, such as the *2007 WEOR*, is based on the Taylor series approximation. DMDC will use this method for *2007 WEOR* variance estimation. In this method a linear approximation to a statistic is formed and then substituted into the formula for calculating the variance of a linear estimate appropriate for the particular sample design. The Taylor series method relies on the simplicity associated with estimating the variance for a linear statistic, even with a complex design, and is valid in large samples.

SUDAAN$^{©}$ is a software package designed to produce variance estimates for complex surveys based on the Taylor series approximation (Research Triangle Institute, Inc., 2004). DMDC will use SUDAAN for variance estimation for the *2007 WEOR*. To facilitate use of SUDAAN, Westat augmented the *2007 WEOR* survey file with variables needed to use SUDAAN (see Table A-3 in A). The added variables are: 1) STATUS (final eligibility indicator); 2) RKWGT0 (the final weight); 3) TVSTR (variance estimation stratum); and TVSTRPOP (total population in variance estimation stratum).

2007 WEOR Response Rates

Response rates are frequently used to measure the quality of a survey. Although the use of response rates as a single measure of the quality of a survey is overstated, they do provide valuable information regarding the success of the survey at representing the population sampled (Madow et al. 1983).

The Council of American Survey Research Organizations (CASRO) has acknowledged that varying operational definitions of response rates can lead to misleading conclusions In an effort to standardize the operational definition and computation of response rates in surveys, CASRO published guidelines and recommendation in 1982 (Council of American Survey Research Organizations, 1982). Beginning in 1995, DMDC standardized its methods for calculating response rates using procedures patterned after those advocated by CASRO. More specifically, the adopted DMDC procedures closely follow CASRO's Sample Type II design.

The main objective of computing response rates is to provide analysts of the *2007 WEOR* data a better understanding on how well the Reserve Component population is represented. To accomplish this goal, response rates are weighted so that they are an estimate of the proportion of the population responding (i.e., response propensity in the population). For example, since the sample was selected with differing sampling rates by sampling strata, the response rates are weighted so each stratum accounts for its appropriate fraction when the total response rate is reported. Observed or unweighted response rates are useful for monitoring the survey during data collection. However, because in most surveys different groups are oversampled, weighted response rates are needed to compare surveys.

Three weighted and unweighted rates will be computed in accordance with the standards defined by CASRO for the *2007 WEOR*. The first rate is the location rate (*LR*) defined as the proportion of eligible sample members that were locatable. The second rate is the completion rate *CR*) defined as the proportion of the located sample that returned usable surveys. The third rate is the response rate (*RR*) computed as the product of the location rate (*LR*) and the completion rate (*CR*), that is:

$$RR = LR \cdot CR.$$

The location, completion and response rates can be also expressed as ratios of the adjusted located sample (N_L), the adjusted eligible sample (N_E), and the usable responses (N_R) as follows:

The *location rate* is defined as

$$LR = \frac{\text{Adjusted located sample}}{\text{Adjusted eligible sample}} = \frac{N_L}{N_E}.$$

The *completion rate* is defined as

$$CR = \frac{\text{Usable responses}}{\text{Adjusted located sample}} = \frac{N_R}{N_L}.$$

The *response rate* is defined as

$$RR = \frac{\text{Usable responses}}{\text{Adjusted eligible sample}} = \frac{N_R}{N_E}.$$

These rates are adjusted for ineligible members to account for the fact that the eligibility of some members is unknown and the proportion of eligibles among the unknowns must be estimated as described in the previous section. The implicit assumption in these calculations is that only ineligible members among the persons with unknown disposition (ELIG = UNK) would proxy- or self-report themselves as ineligible if they return a survey form. That is, the updated frame file is assumed to properly identify all other ineligible members.

To facilitate computation of the CASRO rates and the counts N_L, N_E, and N_R, the variable CAS_ELIG was created to identify the components of *LR*, *CR*, and *RR* as shown in Table 12.

Table 12.
Disposition Codes for CASRO Response Rates (CAS_ELIG)

Eligibility code for CASRO response rates (CAS_ELIG)	Description
ER	Eligible respondent (usable)
ENR_NOQCOMP	Eligible nonrespondent (questionnaire not completed)
ENR_BLANK	Eligible nonrespondent (returned blank questionnaire)
ENR_ACTIVE	Eligible nonrespondent (active refusal)
IN_PR	Proxy-reported ineligible
UNK_NOLOC	Unknown eligibility (nonlocatable member)
UNK_NORET	Unknown eligibility (questionnaire not returned)
IN_FR	Ineligible member in updated frame file

The variable CAS_ELIG was created using the variable SAMP_DC and supplemental survey administration information. The expressions for the numbers of located persons, eligible persons, and usable responses in terms of CAS_ELIG are given below. As notational shorthand, CAS_ELIG codes are used to stand for counts or sum of weights of members in the formulas. For example, ER denotes the count (or sum of weights) of eligible respondents.

The adjusted located sample N_L is defined as the sum of eligible respondents, eligible nonrespondents, and the estimate of members who are assumed to be eligible among the members who did not return the questionnaire. The adjusted located sample N_L, is computed as:

$$N_L = N_{ER} + N_{ENR} + p_E N_{UNK\_NORET}$$

where p_E is the proportion of eligible members observed in the sample computed as:

$$p_E = \frac{N_{ER} + N_{ENR}}{N_{ER} + N_{ENR} + N_{IN\_SR}},$$

and N_{ENR} is the total number of eligible nonrespondents members computed as:

$$N_{ENR} = N_{ENR\_NOQCOMP} + N_{ENR\_BLANK} + N_{ENR\_BLANK}.$$

The adjusted eligible sample N_E is defined as the sum of eligible respondents and the estimate of members who are assumed to be eligible among all members with unknown eligibility. The adjusted eligible sample N_E is computed as:

$$N_E = N_{ER} + N_{ENR} + p_{ER} N_{UNK},$$

where N_{UNK} is the total number of members with unknown eligibility and is computed as

$$N_{UNK} = N_{UNK\_NORET} + N_{UNK\_NOLOC}.$$

The adjusted located count, N_L, and the adjusted eligible count, N_E, can also be expressed by subtracting various counts of ineligible members from the total sample.

The adjusted located count N_L can be computed as

$$N_L = N - N_{IN} - N_{UNK\_NOLOC} - p_{IN\_SR} N_{UNK},$$

where N is the total number of members computed as $N = N_{ER} + N_{IN} + N_{UNK}$, N_{IN} is the total number of ineligible members observed in the sample computed as $N_{IN} = N_{IN\_FR} + N_{IN\_SR}$, and p_{IN_SR} is the proportion of self-reported or proxy reported ineligible members observed in the sample computed as

$$p_{IN\_SR} = \frac{N_{IN\_SR}}{N_{ER} + N_{IN\_SR}} = 1 - p_{ER}.$$

Alternatively, the adjusted eligible count N_E can be computed as

$$N_E = N - N_{IN} - p_{IN\_SR}\ N_{UNK}\ .$$

Response Rate Tables

Table 13 presents both weighted and unweighted location, completion, and response rates for the *2007 WEOR*. Overall, the weighted response rate was 32%. Among the Reserve components, the lowest weighted response rate was obtained for the USMCR (13%) while the highest response rates were realized for the USCGR (44%), USAFR (43%), ANG (42%), and USNR (42%). Significant and historically observed patterns of response rates by paygrade groups were also evident. As paygrade group increases, so does survey response rates. Less variation in response rates was obtained for gender or race/ethnicity groups.

Table 13.

Unweighted and Weighted Location, Completion, and Response Rates for the Full Sample and Categories of Gender, Race/Ethnicity, Reserve Component, Activation Status, and Paygrade Group

Group	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Unweighted			Weighted		
				Location Rate	Completion Rate	Response Rate	Location Rate	Completion Rate	Response Rate
Full Sample	79,369	75,059	23,170	95%	31%	29%	96%	34%	32%
Gender									
Male/Unknown	62,565	59,158	17,660	95%	30%	28%	96%	33%	32%
Female	16,822	15,919	5,510	95%	35%	33%	96%	36%	34%
Race/Ethnicity									
White/Unknown	18,542	17,751	5,610	96%	32%	30%	97%	35%	34%
Black	24,977	23,385	6,680	94%	29%	27%	95%	27%	26%
Hispanic	13,026	12,295	3,491	94%	28%	27%	94%	30%	28%
AIAN	5,983	5,567	1,608	93%	29%	27%	93%	29%	27%
Asian	10,352	9,935	3,484	96%	35%	34%	96%	34%	33%
NHPI	2,871	2,748	954	96%	35%	33%	96%	35%	33%
Two or More Races	3,580	3,351	1,343	94%	40%	38%	94%	40%	38%
Reserve Component									
ARNG	20,158	19,348	4,886	96%	25%	24%	97%	27%	27%
USAR	15,292	14,696	5,014	96%	34%	33%	96%	34%	33%
USNR	10,745	9,395	3,603	87%	38%	34%	90%	46%	42%
USMCR	11,368	10,563	1,285	93%	12%	11%	93%	14%	13%
ARG	9,846	9,591	3,775	97%	39%	38%	98%	43%	42%
USAFR	9,752	9,406	3,670	96%	39%	38%	97%	44%	43%
USCGR	2,081	2,032	937	98%	46%	45%	98%	45%	44%

Table 13. (continued)

Group	Adjusted Eligible Sample	Adjusted Located Sample	Complete Responses	Unweighted Location Rate	Unweighted Completion Rate	Unweighted Response Rate	Weighted Location Rate	Weighted Completion Rate	Weighted Response Rate
Activation Status									
Not active in prior 12 months	58,894	55,682	17,471	95%	31%	30%	96%	33%	32%
Active in prior 12 months	5,572	5,320	1,863	95%	35%	33%	97%	35%	34%
Deactivated in prior 12 months	14,938	14,090	3,836	94%	27%	26%	95%	34%	32%
Paygrade Group									
Unknown	5	5	1	100%	20%	20%	100%	14%	14%
E1-E3	17,664	16,073	1,836	91%	11%	10%	93%	12%	11%
E4	19,755	18,565	3,317	94%	18%	17%	95%	17%	16%
E5-E6	20,063	19,051	6,700	95%	35%	33%	96%	34%	33%
E7-E9	6,577	6,469	3,615	98%	56%	55%	99%	60%	59%
W1-W5	1,253	1,237	728	99%	59%	58%	99%	60%	59%
O1-O3	7,260	7,052	3,235	97%	46%	45%	97%	47%	46%
O4-O6	6,278	6,147	3,738	98%	61%	60%	98%	63%	62%

References

Chromy, J.R. (1987). Design Optimization with Multiple Objectives. *Proceedings of the Survey Research Methods Section.*

Council of American Survey Research Organizations (1982). *On the Definition of Response Rates* (special report of the CASRO task force on completion rates, Lester R. Frankel, Chair). Port Jefferson, NY: Author.

Deever, J. A., & Mason, R. E. (2003, September). DMDC Sample Planning Tool (Version 2.1) [Computer software]. Arlington, VA: DMDC.

DMDC. (2008). *2007 Workplace and Equal Opportunity Survey of Reserve Component Members: Administration, Datasets, and Codebook* (Report No. 2007-040). Arlington, VA: Author.

Kalton, G., & Kasprzyk, D. (1989). The Treatment of Missing Survey Data. *Survey Methodology, 12,* 1–16.

Kass, G. (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics*, 29, 119-127.

Kish, L. (1992). Weighting for Unequal Pi. *Journal of Official Statistics, 8,* 183-200.

Madow, W., Nisselson, H., & Olkin, I. (1983). *Incomplete Data on Sample Surveys, Volume 1 – Report and Case Studies.* New York: Academic Press.

Mason, R.E., Kavee, J.A.,Wheeless, S.C., George, B.J., Riemer, R.A., & Elig, T.W. (1996). *The 1995 Armed Forces Sexual Harassment Survey: Statistical Methodology Report* (Report No. 96-016). Arlington VA: Defense Manpower Data Center (DTIC No. AD A323 943).

Riemer, R. A. and Kroeger, K. R. (2002). *Statistical design of the Status of Forces Surveys of Active-Duty Members* (Report No. 2002-033). Arlington, VA: DMDC.

Research Triangle Institute, Inc. (2004). SUDAAN© Language Manual, Release 9.0. Research Triangle Park, NC: Research Triangle Institute.

Skinner, C., Holt, D., & Smith, T. (Eds) (1989). *Analysis of Complex Surveys.* New York: Wiley.

Appendix A.
Supplementary Tables

Supplementary Tables

Table A-1.
Allocation Solution for Reporting Domains

Domain	Label	Population	95% HWCI	Allocation	Estimated n	Pct Sampled	Design Effect
1	All Domains	805,144	0.01	22,870	82,128	10.32	3.59
2	DOD	797,241	0.01	22,107	79,966	10.15	3.53
3	US Coast Guard	7,903	0.04	763	2,162	27.69	1.29
4	NotASOC12	605,057	0.02	17,069	61,458	10.28	4.41
5	ActSOC12	200,087	0.03	5,801	20,672	10.45	6.24
6	Army National Guard	339,644	0.02	5,359	20,476	6.07	2.41
7	US Army Reserve	179,401	0.02	4,670	15,504	8.73	3.01
8	US Naval Reserve	66,389	0.03	3,751	11,718	18.59	4.48
9	US Marine Corps Reserve	37,809	0.04	1,987	12,193	32.57	3.05
10	Air National Guard	103,379	0.04	3,312	9,900	9.66	4.50
11	US Air Force Reserve	70,619	0.04	3,030	9,907	14.15	4.28
12	Enlisted	684,506	0.01	16,412	67,255	9.94	3.42
13	Officer	120,626	0.02	6,457	14,871	12.47	1.87
14	E1-E4	318,477	0.02	6,942	39,651	12.54	2.88
15	E5-E9	366,041	0.02	9,472	27,678	7.68	3.93
16	W1-W5	9,797	0.05	599	1,242	12.99	1.55
17	O1-O3	46,012	0.02	2,928	7,321	16.07	1.76
18	O4-O6	64,817	0.03	2,931	6,307	9.83	1.97
19	Male	665,055	0.01	17,881	64,625	9.82	3.49
20	Female	140,089	0.03	4,989	17,488	12.68	3.98
21	Hispanic	71,971	0.02	3,044	13,567	19.05	1.96
22	Hispanic*ARNG	26,156	0.05	588	2,589	9.96	1.52
23	Hispanic*USAR	21,369	0.05	657	2,313	10.91	1.70
24	Hispanic*USNR	7,403	0.05	459	1,673	23.63	1.19
25	Hispanic*USMCR	4,705	0.05	403	4,025	86.30	1.07
26	Hispanic*ANG	6,342	0.05	421	1,303	20.69	1.09
27	Hispanic*USAFR	5,303	0.05	453	1,449	27.51	1.17
28	Hispanic*Enlisted	65,903	0.03	2,131	11,396	17.48	1.62
29	Hispanic*Officers	6,065	0.03	914	2,170	36.15	1.00
30	Hispanic*E1-E4	32,134	0.04	795	6,866	21.52	1.51
31	Hispanic*E5-E9	33,772	0.03	1,335	4,541	13.64	1.57
32	Hispanic*O1-O3	2,869	0.05	400	1,074	37.73	1.03
33	Hispanic*O4-O6	2,700	0.05	415	862	32.20	1.07

Table A-1. (continued)

Domain	Label	Population	95% HWCI	Allocation	Estimated n	Pct Sampled	Design Effect
34	White	582,687	0.02	5,747	18,917	3.28	1.56
35	White*ARNG	259,713	0.03	2,006	7,084	2.74	1.39
36	White*USAR	108,895	0.04	1,105	3,007	2.79	1.60
37	White*USNR	43,988	0.05	559	1,215	2.90	1.44
38	White*USMCR	28,269	0.05	431	3,476	12.41	1.12
39	White*ANG	84,065	0.04	673	1,601	1.92	1.36
40	White*USAFR	51,560	0.05	569	1,287	2.52	1.47
41	White*Enlisted	483,678	0.02	3,164	13,875	2.90	1.19
42	White*Officers	99,003	0.02	2,583	5,042	5.15	1.07
43	White*E1-E4	223,875	0.03	1,570	9,963	4.48	1.15
44	White*E5-E9	259,809	0.03	1,594	3,940	1.54	1.22
45	White*O1-O3	35,834	0.03	1,135	2,497	7.03	1.06
46	White*O4-O6	54,828	0.03	1,154	2,024	3.73	1.06
47	Black	118,250	0.02	6,230	26,064	22.43	2.60
48	Black*ARNG	44,703	0.04	900	4,961	11.20	1.44
49	Black*USAR	40,871	0.04	1,271	4,733	11.75	1.77
50	Black*USNR	9,838	0.03	1,213	4,823	53.14	1.13
51	Black*USMCR	2,894	0.07	589	2,764	97.02	3.34
52	Black*ANG	8,605	0.03	1,049	3,802	44.75	0.98
53	Black*USAFR	10,984	0.03	1,177	4,669	43.06	1.10
54	Black*Enlisted	106,891	0.02	4,759	22,137	21.09	2.39
55	Black*Officers	11,357	0.03	1,472	3,931	35.11	1.06
56	Black*E1-E4	47,643	0.04	1,350	11,793	25.03	2.06
57	Black*E5-E9	59,250	0.02	3,409	10,368	17.92	2.21
58	Black*O1-O3	5,166	0.04	630	1,833	35.92	1.05
59	Black*O4-O6	5,422	0.04	687	1,734	32.42	1.14
60	AIAN	6,257	0.02	2,607	6,132	100.00	1.52
61	AIAN*ARNG	2,814	0.04	1,172	2,758	100.00	2.32
62	AIAN*USAR	1,108	0.08	462	1,086	100.00	2.72
63	AIAN*USNR	1,011	0.08	421	991	100.00	2.74
64	AIAN*USMCR	240	0.17	100	235	100.00	2.92
65	AIAN*ANG	703	0.10	293	689	100.00	2.82
66	AIAN*USAFR	297	0.15	124	291	100.00	2.92
67	AIAN*Enlisted	5,727	0.03	2,386	5,612	100.00	1.64
68	AIAN*Officers	530	0.11	221	519	100.00	2.86

Table A-1. (continued)

Domain	Label	Population	95% HWCI	Allocation	Estimated n	Pct Sampled	Design Effect
69	Asian	19,272	0.02	2,915	10,621	55.91	1.27
70	Asian*ARNG	6,258	0.04	693	3,037	48.87	1.15
71	Asian*USAR	5,484	0.04	754	2,690	49.51	1.25
72	Asian*USNR	2,511	0.05	449	1,359	57.25	1.16
73	Asian*USMCR	1,261	0.08	309	1,246	99.76	2.32
74	Asian*ANG	2,322	0.05	381	1,152	50.00	0.99
75	Asian*USAFR	1,401	0.05	326	1,125	81.01	0.91
76	Asian*Enlisted	16,490	0.02	1,973	8,310	51.13	1.13
77	Asian*Officers	2,782	0.03	943	2,312	84.20	0.96
78	NHPI	2,942	0.03	738	2,904	100.00	0.77
79	NHPI*USAR	1,674	0.05	420	1,652	100.00	1.10
80	NHPI*USNR	237	0.15	59	234	100.00	1.45
81	NHPI*USMCR	184	0.18	46	182	100.00	1.47
82	NHPI*ANG	411	0.12	103	406	100.00	1.42
83	NHPI*USAFR	419	0.11	105	414	100.00	1.42
84	NHPI*Enlisted	2,647	0.04	664	2,613	100.00	0.85
85	NHPI*Officers	295	0.14	74	291	100.00	1.45
86	MultiRace	3,765	0.03	1,587	3,742	100.00	1.27
87	MultiRace*USNR	1,401	0.06	591	1,393	100.00	2.06
88	MultiRace*USMCR	256	0.15	108	254	100.00	2.45
89	MultiRace*ANG	931	0.07	392	925	100.00	2.22
90	MultiRace*USAFR	655	0.09	276	651	100.00	2.31
91	MultiRace*Enlisted	3,170	0.03	1,336	3,151	100.00	1.47
92	MultiRace*Officers	594	0.09	250	590	100.00	2.33
93	Northern U.S.	299,121	0.03	5,787	21,249	7.18	4.34
94	Southern U.S.	323,943	0.02	9,574	34,586	10.81	5.54
95	Western U.S.	154,703	0.04	6,288	21,646	14.17	8.50

Table A-2.

Summary of Weighting Adjustment Variability

Nonresponse Adjustment Factor or Weight	Response Status (STATUS)	Minimum	Mean	Maximum	Median	Standard Deviation	Coefficient of Variation	Skewness
Unknown eligibility adjustment (NRFACT_UNK00)	Respondents (1)	1.3250	2.8228	9.9265	2.2237	1.6514	58.5018	2.2100
Unknown eligibility adjustment (NRFACT_UNK00)	Nonrespondents (2)	1.3250	3.4925	9.9265	2.4322	2.0826	59.6289	1.3800
Unknown eligibility adjustment (NRFACT_UNK00)	Inelligible (3)	1.3250	3.1801	9.9265	2.4322	1.7412	54.7530	1.9709
Unknown eligibility adjustment (NRFACT_UNK00)	Nonrespondents of unknown eligibility (4)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Unknown eligibility adjusted weight (NRWGT_UNK00)	Respondents (1)	1.5229	28.8573	271.4235	6.9768	50.7930	176.0142	2.5637
Unknown eligibility adjusted weight (NRWGT_UNK00)	Nonrespondents (2)	1.5229	33.1471	271.4235	7.7597	53.7182	162.0598	2.2530
Unknown eligibility adjusted weight (NRWGT_UNK00)	Inelligible (3)	1.5229	21.2092	271.4235	5.7748	39.9420	188.3238	3.2123
Unknown eligibility adjusted weight (NRWGT_UNK00)	Nonrespondents of unknown eligibility (4)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Eligible Nonresponse Adjustment (NRFACT_NR00)	Respondents (1)	1.0000	1.1183	1.4049	1.1043	0.0694	6.2101	1.0268
Eligible Nonresponse Adjustment (NRFACT_NR00)	Nonrespondents (2)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Eligible Nonresponse Adjustment (NRFACT_NR00)	Inelligible (3)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Eligible Nonresponse Adjusted Weight (NRWGT_NR00)	Respondents (1)	1.5229	32.7800	306.1677	7.7412	58.0387	177.0550	2.5630
Eligible Nonresponse Adjusted Weight (NRWGT_NR00)	Nonrespondents (2)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Eligible Nonresponse Adjusted Weight (NRWGT_NR00)	Inelligible (3)	1.5229	21.2092	271.4235	5.7748	39.9420	188.3238	3.2123
Eligible Nonresponse Adjusted Weight (NRWGT_NR00)	Nonrespondents of unknown eligibility (4)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table A-3.
Collapsed Design Strata Used for Variance Estimation in SUDAAN

Variance Strata (TVSTR)	Total Population in Variance Strata (TVSTRPOP)	Achieved Sample Size	Design Strata (STRAT)
1	11,371	1,096	1
2	9,441	577	2
3	972	365	3
4	540	146	4
6	3,574	294	5, 6
7	8,284	3,683	7
8	93,706	1,042	8
9	12,912	893	9
10	9,230	297	10
11	17,930	600	11
12	9,228	106	12
14	2,586	161	13, 14
15	16,887	2,157	15
16	14,964	989	16
17	1,163	372	17
18	622	134	18
19	6,650	767	19
20	3,510	228	20
21	606	214	21
22	2,857	1,404	22
23	1,796	623	23
24	621	542	24
25	622	336	25
26	292	119	26
27	6,858	679	27
28	7,858	503	28
29	745	265	29
30	726	189	30

Table A-3. (continued)

Variance Strata (TVSTR)	Total Population in Variance Strata (TVSTRPOP)	Achieved Sample Size	Design Strata (STRAT)
31	2,333	222	31
32	2,025	131	32
33	448	175	33
34	31,426	1,022	34
35	36,494	393	35
36	6,823	457	36
37	9,934	311	37
38	7,732	249	38
39	6,985	79	39
40	2,226	166	40
41	2,120	71	41
42	8,079	978	42
43	14,261	901	43
44	1,344	413	44
45	1,750	361	45
46	5,263	569	46
47	7,127	440	47
48	1,117	364	48
49	1,002	213	49
50	1,931	899	50
51	1,506	500	51
52	873	730	52
53	539	273	53
54	479	181	54
55	2,219	603	55
56	2,921	552	56
57	285	113	57
58	393	115	58
59	784	206	59
60	789	155	60

Table A-3. (continued)

Variance Strata (TVSTR)	Total Population in Variance Strata (TVSTRPOP)	Achieved Sample Size	Design Strata (STRAT)
61	6,390	244	61
62	18,804	323	62
63	2,434	168	63
64	8,581	285	64
66	5,541	141	65, 66
68	1,949	92	67, 68
69	2,241	1,663	69
70	4,013	1,720	70
71	771	355	71
72	1,131	768	72
73	1,661	710	73
74	604	359	74
75	1,054	459	75
76	444	382	76
77	402	232	77
78	2,994	2,994	78
79	1,377	931	79
81	4,711	379	81
83	1,412	115	82, 83
84	20,120	2,849	80, 84
85	1,738	1,738	85
86	901	901	86
87	1,190	1,190	87
88	1,069	272	88
89	3,664	632	89
90	509	193	90

Table A-3. (continued)

Variance Strata (TVSTR)	Total Population in Variance Strata (TVSTRPOP)	Achieved Sample Size	Design Strata (STRAT)
92	1,100	215	91, 92
93	14,817	442	93
94	45,398	458	94
95	3,368	218	95
96	6,625	206	96
98	12,014	197	97, 98
100	1,843	93	99, 100
101	1,559	1,027	101
102	3,969	1,480	102
103	650	279	103
104	778	464	104
105	1,649	601	105
106	479	269	106
107	1,207	486	107
108	244	198	108
109	392	208	109
110	904	285	110
111	2,353	505	111
112	370	144	112
113	369	112	113
114	605	132	114
115	6,613	222	115
117	1,638	108	117
118	4,892	154	118
120	6,539	131	119, 120

Table A-3. (continued)

Variance Strata (TVSTR)	Total Population in Variance Strata (TVSTRPOP)	Achieved Sample Size	Design Strata (STRAT)
122	1,614	79	121, 122
123	1,818	1,030	123
124	4,035	1,277	124
125	289	123	125
126	316	111	126
127	1,226	632	127
128	1,918	591	128
129	1,167	901	129
130	6,257	6,257	130
131	2,942	2,942	131
132	3,765	3,765	132
133	1,377	672	133
134	5,837	346	134
135	2,361	155	135
136	412	412	136
137	1,194	478	137
138	2,124	1,485	138
139	3,183	158	139
140	1,926	163	140
141	24,535	330	116, 141
142	5,729	273	142
143	1,083	307	143
144	2,017	404	144
145	2,544	509	145
146	801	174	146
147	835	171	147

REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.